

Downscaling SMAP soil moisture using a hybrid machine-learning algorithm

Abhilash Singh^{a,*}, M. Niranjannaik^{a,b}, Kumar Gaurav^{a,*}

^a FGRS Laboratory, Department of Earth and Environmental Sciences, IISER Bhopal, Bypass road, Bhopal, India

^b Department of Earth Sciences, Indian Institute of Technology, Gandhinagar, India

HIGHLIGHTS

- Novel downscaling approach integrates Scorpion Hunting Strategy and fuzzy inference for SMAP.
- The algorithm was validated with in-situ data from Bhopal and Kosi Fan, demonstrating strong accuracy.
- Downscaled maps correlate well with in-situ data and capture fine-scale soil moisture variability.
- Broad applicability for agriculture, hydrology, and environmental monitoring is highlighted.

ARTICLE INFO

Keywords:

Soil moisture
Downscaling
SMAP
Nature-inspired algorithm
Scorpion hunting strategy
Fuzzy inference system

ABSTRACT

This study introduces an innovative approach for downscaling Soil Moisture Active Passive (SMAP) satellite soil moisture data from a coarse spatial resolution of 9 km to a refined resolution of 1 km. The proposed method uniquely integrates the Scorpion Hunting Strategy Algorithm, a novel nature-inspired optimisation technique, with a fuzzy inference system. Validation was performed using field measurements obtained from two climatically distinct and data-scarce regions: Bhopal (semi-arid) and the Kosi Fan (humid sub-tropical). The resulting high-resolution soil moisture maps demonstrated a strong correlations with the in-situ observations (correlation coefficients (R) of 0.79 and 0.76, and root mean square errors (RMSE) of 0.053 m³/m³ and 0.043 m³/m³ for Bhopal and Kosi Fan, respectively). These findings point to the robustness and versatility of the proposed approach in capturing intricate spatial variations of soil moisture across diverse climatic regimes. Performance of the algorithm was further verified by testing with publicly available data from other fields, enhancing its broader applicability. Overall, this novel approach contributes a significant advancement in soil moisture downscaling techniques, offering valuable insights and practical benefits for application in agriculture, hydrological forecasting, and environmental monitoring.

1. Introduction

Soil moisture is a key environmental parameter that has a substantial influence on hydrological processes, agricultural production, and ecological stability [1,2]. Accurate, high-resolution soil moisture data are essential for effective water resource management, drought monitoring, irrigation scheduling, and climate resilience planning. The Soil Moisture Active Passive (SMAP) mission provides valuable global-scale soil moisture observations; however, its relatively coarse spatial resolution (9 km) limits its applicability to localised management and decision-making processes [3]. To address this, spatial downscaling

techniques have been developed. Advanced machine learning models have proven effective in downscaling data from coarse to finer spatial resolutions [4].

Current advances in downscaling soil moisture are mainly based on conventional machine learning methods such as Random Forest, Neural Networks, and Support Vector Regression, with varying degrees of success depending on the climatic and geographical conditions [5–7]. However, the integration of nature-inspired algorithms/computation with fuzzy theory is a promising but largely unexplored. The hybrid models are more effective in data-scarce regions [8].

* Corresponding authors.

Email addresses: a.singh4@leeds.ac.uk, abhilash.iiserb@gmail.com (A. Singh), niranjannaik@iiserb.ac.in (M. Niranjannaik), kgaurav@iiserb.ac.in (K. Gaurav).

<https://doi.org/10.1016/j.asoc.2026.114766>

Received 24 May 2025; Received in revised form 27 November 2025; Accepted 28 January 2026

Available online 6 February 2026

1568-4946/© 2026 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

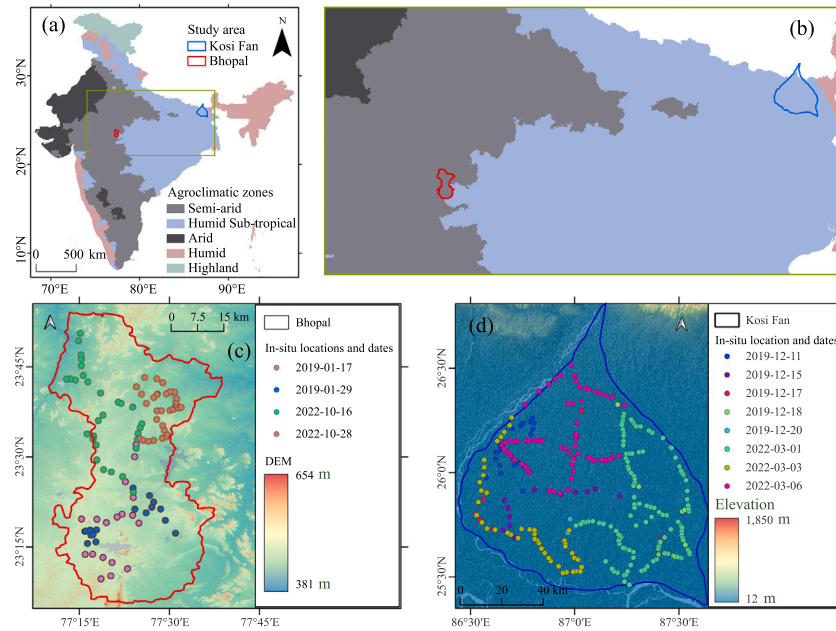


Fig. 1. (a & b) The map in the upper panel displays the Köppen-Geiger climate classification for India. The red and blue polygons highlight the Bhopal and Kosi Fan study areas, respectively. (c) In-situ measurement sites at Bhopal: Measurements were taken at 99 locations, with some grid areas having multiple measurements due to their heterogeneity. To determine a representative value, we calculated the average of these measurements, resulting in a total of 91 effective measurements. (d) In-situ measurement sites at Kosi Fan: A total of 307 measurement sites were surveyed, with 281 effective measurements used for analysis.

The Scorpion Hunting Strategy (SHS), a novel nature-inspired optimisation algorithm, has demonstrated significant potential for solving complex spatial problems efficiently [9]. Coupling SHS with fuzzy inference systems (FIS) may further enhance downscaling precision by effectively managing data uncertainties and incorporating human-like reasoning.

This study introduces a novel soil moisture downscaling approach by integrating the SHS algorithm with fuzzy inference systems to improve the spatial resolution of SMAP soil moisture products from 9 km to 1 km. The robustness and adaptability of the model are validated across climatically diverse and data-scarce regions, specifically Bhopal (a semi-arid zone) and the Kosi Fan (a humid sub-tropical zone) in India. This study addresses the following research questions:

- How effectively can the Scorpion Hunting Strategy algorithm combined with fuzzy inference systems downscale SMAP soil moisture data from 9 km to 1 km resolution?
- Can the integrated SHS-FIS approach generalise effectively across varied climatic and geographic contexts, thus providing broader applicability for agricultural, hydrological, and environmental monitoring purposes?
- How does the performance of the proposed approach compare to conventional and other advanced machine learning downscaling methods reported in recent literature?
- How does the performance of the proposed algorithm compare when applied to publicly available datasets outside of the hydrology domain?

The rest of the manuscript is divided into three sections. Section 2 discusses the study areas and the datasets employed in this study, followed by the data pre-processing and feature engineering steps. Section 3 details the development of the downscaling algorithm. Section 4 presents and discusses the results obtained by applying the proposed downscaling algorithm to different climate settings. Finally, Section 5 summarises the main findings of this study.

2. Study area, datasets and pre-processing

2.1. Study area

We have conducted this study in two agriculture dominated regions with diverse climates in India (Fig. 1(a) and (b)) [10–12]. Bhopal is the capital city of Madhya Pradesh in central India. It lies in a semi-arid climate zone (Fig. 1(c)). We conducted two field campaigns in the Bhopal region to obtain in-situ soil moisture measurements using a calibrated time domain reflectometer (TDR). The first campaign took place from January 17–29, 2019, and the second from October 16–28, 2022. We collected surface soil moisture measurements at 91 different locations (Fig. 1(c)). These locations were chosen using a random grid sampling approach, which is illustrated in Fig. 2(a). Along with the moisture readings, we recorded the coordinates of each location. The coordinates are captured using a Garmin GPSmap 64s device (Fig. 2(b)).

The Kosi Fan, located in northern Bihar within the Himalayan foreland, lies in the humid sub-tropical climate zone (Fig. 1(d)). We conducted field campaigns in the Kosi Fan from December 11–20, 2019, and March 1–6, 2022. We collected 281 in-situ soil moisture measurements, which are well distributed throughout the region [13].

These two regions exhibit distinct characteristics in terms of climate (flood-prone and drought-prone), land use, geology (soil and grain size), and topography (flat and uneven), encompassing a wide range of variations in each feature. Therefore, these field sites are ideal for assessing the downscaling of soil moisture under heterogeneous conditions.

2.2. Datasets

We used a combination of optical, radar, and radiometer datasets from various satellites. Additionally, we conducted extensive field surveys to collect in-situ soil moisture measurements during the satellite pass of Sentinel-1.

2.2.1. SMAP

We accessed SMAP data from the National Snow and Ice Data Center (NSIDC) (<https://nsidc.org/data/smap/data>). It includes an L-band

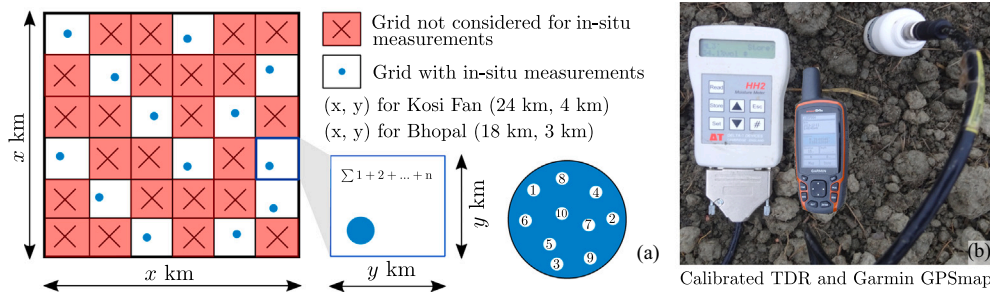


Fig. 2. (a) Schematic representation of the random grid sampling approach used to collect in-situ measurements at both study sites. (b) Soil-specific calibrated TDR sensor for surface soil moisture measurement and handheld GPS used in the study.

passive radiometer with a frequency of 1.41 GHz and an active radar with a frequency of 1.26 GHz. The passive radiometers measure brightness temperatures in the microwave region, while the radar uses SAR to measure transmitted pulses from the sensor and reflected microwave pulses from the Earth's surface at high spatial resolution. One of the instruments on SMAP (radar) stopped transmitting data from July 7, 2015. Despite this loss, the SMAP mission continues to function and provides soil moisture data using the radiometer instrument of level-2 passive soil moisture (L2_SM_P) at a 36 km resolution. Additionally, the downscaled product of (L2_SM_P) is level-2 passive enhanced soil moisture (L2_SM_PE), available at a spatial resolution of 9 km. Furthermore, another level-2 soil moisture product (L2_SM_SP), combining active sensors from Sentinel-1 and passive radiometers, is also accessible.

The SMAP L-band SAR offers high resolution backscatter measurements (1 and 3 km) compared to the coarser radiometer measurements. However, radar sensitivity to surface roughness and vegetation scattering limits its soil moisture sensing accuracy [14]. The SMAP combined with L-band radar and radiometer measurements provides higher accuracy and resolution (9 km) of soil moisture [15].

For our analysis, we opted for the daily composite level-3 SMAP product L3_SM_P_E (9 km) dataset [16], which is a daily composite of global SMAP Level-2 L2_SM_P_E data. We obtained 12 SMAP L3_SM_P_E datasets (8 on the Kosi Fan and 4 on the Bhopal) corresponding to the in-situ measurements taken on the aforementioned dates.

2.2.2. Sentinel-1/2

To support the downscaling of coarse-resolution SMAP soil moisture data, high-resolution remote sensing inputs are acquired from the Copernicus Open Access Hub (<https://scihub.copernicus.eu/>). Specifically, images from the Sentinel-1 and Sentinel-2 missions, operated by the European Space Agency (ESA), were utilised due to its complementary sensing capabilities and spatial detail.

The Sentinel-1 mission comprises Sentinel-1A (launched in 2014) and Sentinel-1B (launched in 2016), both operating in the C-band at 5.405 GHz as part of ESA's Copernicus Programme. These satellites provide global coverage with a revisit period of six days and a spatial resolution of 10×10 m. Sentinel-1 supports dual polarisation (VV and VH) and offers four acquisition modes: Strip Map (SM), Interferometric Wide Swath (IW), Extra-Wide Swath (EW), and Wave Mode. For this study, Level-1 Ground Range Detected (GRD) products acquired in IW mode were employed to capture spatial and temporal variability in surface roughness and vegetation, which are crucial for soil moisture estimation.

On top of this, optical data from Sentinel-2A and Sentinel-2B, which were launched in June 2015 and March 2017, respectively, were also included. These satellites have high-resolution multispectral images at 10 m spatial resolution with a five-day revisit cycle. Their fine-scale resolution and spectral richness provide rich information regarding land surface characteristics, enhancing the accuracy of the downscaling procedure for soil moisture.

2.2.3. MODIS

To further enhance the downscaling of coarse-resolution SMAP soil moisture observations, we incorporated land surface temperature (LST) data from the Moderate Resolution Imaging Spectroradiometer (MODIS), onboard NASA's Terra and Aqua satellites, both polar-orbiting and sun-synchronous platforms [17]. Daily LST products are sourced from the official MODIS data portal (<https://modis.gsfc.nasa.gov/>), offering consistent and long-term thermal information at multiple spatial resolutions, including 1×1 km and 6×6 km grids.

The MODIS LST dataset is derived from thermal infrared (TIR) measurements captured in Band 31 ($10.78\text{--}11.28 \mu\text{m}$) and Band 32 ($11.77\text{--}12.27 \mu\text{m}$), providing reliable indicators of surface energy balance, soil moisture variability, and vegetation-temperature interactions. These properties are particularly valuable for downscaling applications, as LST is closely linked to surface moisture dynamics and can help refine spatial heterogeneity in SMAP-derived soil moisture estimates.

2.3. Pre-processing

To facilitate the downscaling of coarse-resolution SMAP soil moisture data, a structured pre-processing workflow is implemented for Sentinel-1 and Sentinel-2 data. The primary objective was to ensure spatial consistency and extract relevant physical variables that influence soil moisture variability at finer scales. We first reprojected the EASE SMAP product to UTM Zone 43N for Bhopal and UTM Zone 45N for the Kosi Fan. This reprojection process was consistently applied to all other datasets to ensure spatial alignment and compatibility.

Sentinel-1 Level-1 GRD images were processed using the Sentinel Application Platform (SNAP v8.0). The pre-processing steps included radiometric calibration to convert raw digital numbers into radar backscatter values, followed by multi-looking with a factor of six to reduce speckle noise and improve radiometric accuracy. A refined Lee filter is applied to further suppress speckle while preserving important textural features. Geometric terrain correction is then performed using the Shuttle Radar Topography Mission (SRTM) digital elevation model at a spatial resolution of 30 m, ensuring accurate geolocation and topographic normalisation. The output consisted of radar backscattering coefficients in linear scale for both VV and VH polarisations (σ_{VV}^0 and σ_{VH}^0), along with the local incidence angle (θ), all resampled to a spatial resolution of 60 m. These radar-derived parameters provide critical insights into surface roughness and vegetation structure, which are key controls on soil moisture distribution.

Level-2A orthorectified and atmospherically corrected surface reflectance products of Sentinel-2 were utilised for the optical domain. Images are selected based on temporal alignment with field campaigns and a cloud cover threshold of less than 10%. Cloud-contaminated pixels are masked, and the median surface reflectance is computed across available scenes within each campaign period. The Normalised Difference Vegetation Index (NDVI) is derived using the near-infrared (B8) and red (B4) bands, providing an indicator of vegetation vigour and

Table 1
Details of the satellite data used in this study.

Satellite	Spatial resolution	Sensor type	Temporal resolution	Band details
Sentinel-2	10 m	Optical	Five days	646–685 nm & 774–907 nm
Sentinel-1	10 m	Radar	Six days	C- band (5.404 GHz)
SMAP	9 km	Active and Passive microwave	2–3 days	L-band (1.41 GHz)
MODIS LST	1 & 6 km	Thermal	Daily	10.78–11.28, 11.77–12.27, 3.93–3.99, and 4.02–4.08 μm

density, which strongly influences evapotranspiration and soil moisture retention.

To ensure interoperability among datasets and alignment with the SMAP soil moisture product, all variables derived from Sentinel-1 and Sentinel-2 are resampled to spatial resolutions of 1 km and 9 km using bilinear interpolation. We used the 1 km resolution data as input for training and prediction in the downscaling framework, while the 9 km data represented the original SMAP grid. This harmonised, multi-resolution dataset structure enabled a supervised learning approach to infer high-resolution soil moisture estimates with improved spatial fidelity (Table 1).

3. Development of the downscaling algorithm

We leverage the SHS algorithm, together with a Sugeno-type FIS, to develop the proposed downscaling framework for SMAP soil moisture products. The SHS algorithm, introduced by Singh et al. [9] and illustrated in Fig. 3, mimics the behavioural dynamics of scorpions during prey capture and intra-species encounters. In SHS, each candidate solution is treated as a “scorpion”, and the search process evolves through three biologically inspired variation operators: alpha-mutation, beta-mutation, and alpha-beta hybrid mutation, applied depending on whether a scorpion encounters prey (better solution), a predator (worse solution), or a peer (another candidate). These operators are parameterised through claw power (ω) and sting power (ψ), which together determine the adaptive step size and directional bias in the optimisation space. Despite having multiple eyes, real scorpions predominantly rely on tactile sensing, claws, and sting to subdue prey, an analogy captured in SHS through selective pressure mechanisms in which stronger

scorpions (solutions with lower cost) prevail. In our implementation, SHS uses a population of 10 candidate solutions, random uniform initialisation within bounded parameter ranges (± 10), mutation rate $\mu = 0.2$ with damping $\mu_{\text{damp}} = 0.98$, and a fixed maximum of 1000 iterations.

To utilise the potential of nature-inspired algorithms in solving the spatial downscaling problem, we integrated the SHS algorithm with the FIS of type Sugeno using fuzzy C-mean (FCM) clustering [9]. From here onwards, we refer to this model as scorpion downscaling algorithm. To determine the optimal number of clusters for the Sugeno-type FIS, we employed a two-step approach that considered both data characteristics and domain knowledge (building upon the principles outlined in Sugeno 1985 [18]). First, we utilised the Elbow method to identify an appropriate cluster size. This method analyses the trade-off between the number of clusters and the explained variance. We identified five clusters as the optimal configuration for this application. Next, we initialised the cluster centers randomly and employed Gaussian membership functions (gaussmf) to quantify the degree of membership for each data point within a cluster. These membership functions are defined by range and variance parameters. An iterative optimisation process followed, where the SHS algorithm was utilised to update the membership function parameters for each cluster (as shown in Fig. 4) to achieve a well-defined clustering structure. The optimisation aims to minimise the fuzzy-model cost function, which evaluates the prediction error between FIS outputs and SMAP soil moisture using all training samples. For reproducibility, all experiments were run five times with different random seeds, and the best-cost solution was retained to construct the final SHS-optimised FIS.

Following the convergence of the membership function parameters, each data point is assigned to the cluster with the highest degree of membership. This process effectively partitions the data into meaningful groups based on their characteristics. Next, we define the input and output variables for the FIS. We leverage the identified input features and target SMAP soil moisture level for these roles, respectively. We establish linguistic variables for each input and output to capture the relationships between these variables. These linguistic variables represent the fuzzy nature of the data, allowing for the creation of fuzzy rules (presented in Eq. (1)). These rules express the relationships between the input features and the predicted soil moisture values. We have one rule for each fuzzy cluster.

$$\text{Input}_1, \text{cluster}(j) \& \text{Input}_2, \text{cluster}(j) \& \dots \& \text{Input}_f, \text{cluster}(j) \xrightarrow{\text{Sugeno}} \text{Output cluster}(j) \quad (1)$$

where j ranges from 1 to 5, and f represents the number of input features. In total, five fuzzy rules are established, where each rule employs the AND fuzzy logic operator to map an input with f features to the response variable. These rules connect to specific outcomes, expressed using linguistic terms from the output variable's domain. The overall process is summarised in Fig. 5 and Algorithm 1.

We now use the proposed novel algorithm to downscale the SMAP soil moisture information. For this, we considered both the study area (i.e., Bhopal and Kosi Fan) and downscaled the 9 km SMAP soil moisture information to 1 km soil moisture using Sentinel-1 backscatters (VV and VH), NDVI and LST as the potential inputs [19]. In doing so, we extracted these input features at a spatial resolution of 9 km along with the

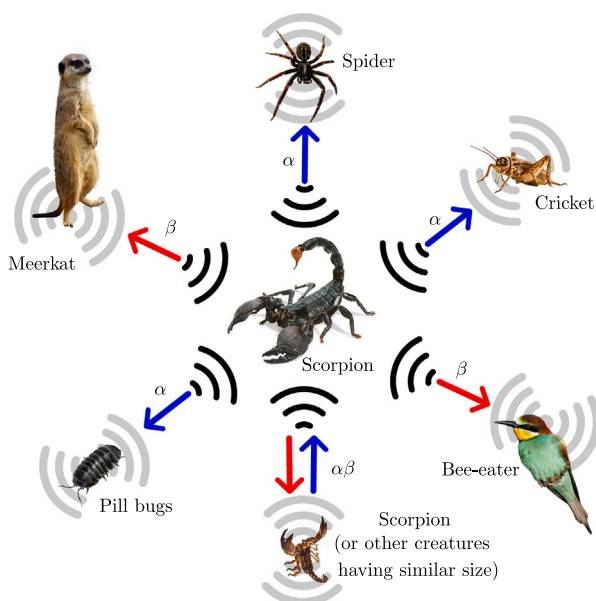


Fig. 3. Schematic showing how the α and β operators behave when a scorpion encounters prey or a hunter, respectively [9].

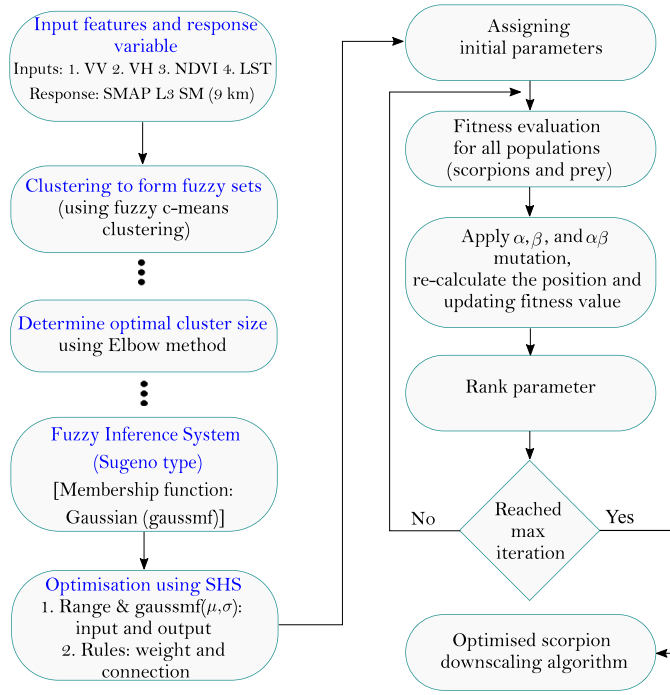


Fig. 4. Flowchart illustrating the working of the scorpion downscaling algorithm.

corresponding SMAP soil moisture, which serves as the response variable in the model. We trained the Scorpion downscaling algorithm using the datasets at 9 km. Once the model is trained, it is assumed that the dynamics of the mapping function between the input features and response variable will be independent of spatial scale [20,21]. Hence, we fed finer scale (1 km) input features into the model to generate downscaled soil moisture at a spatial resolution of 1 km.

The computational complexity of the SHS-FIS optimisation framework is primarily governed by the SHS population updates and repeated cost-function evaluations. For a population of n_{pop} candidate solutions, n_{iter} iterations, and N training samples, the overall complexity is approximately $\mathcal{O}(n_{pop} \times n_{iter} \times N)$. In our configuration ($n_{pop} = 10$, $n_{iter} = 1000$), this results in a lightweight optimisation cycle. Empirically, complete training of the SHS-FIS model required 18–22 s per run on an NVIDIA A40 GPU (46 GB VRAM). These runtimes correspond to a single optimisation pass, and all experiments were repeated five times using different random seeds. The modest computational footprint demonstrates that the proposed downscaling framework is computationally efficient and suitable even for low-resource environments.

4. Results and discussion

4.1. Model performance

Figs. 6 and 7 depict the spatial soil moisture distributions at native SMAP resolution (9 km) and downscaled resolution (1 km) for the Bhopal and the Kosi Fan study regions, respectively. The visual comparison highlights the ability of the proposed algorithm to successfully capture fine-scale spatial heterogeneity that is otherwise obscured in the coarser SMAP product. For Bhopal, the downscaled soil moisture values ranged from 0.17 to 0.21 m^3/m^3 , while for the Kosi Fan, the values varied more widely between 0.22 and 0.40 m^3/m^3 . These enhanced spatial patterns demonstrate the model's sensitivity to local surface characteristics such as vegetation, roughness, and temperature, as derived from high-resolution remote sensing inputs.

Quantitative validation is conducted by comparing the downscaled soil moisture against in-situ field measurements. For the Bhopal site,

Algorithm 1 Pseudo code for scorpion downscaling algorithm.

```

1: Inputs: Input features and response variable as data (data.Train: 70% and data.Test: 30%)  $\triangleright$  For both the study sites
2: Inputs: In-situ soil moisture as in-situ_SM  $\triangleright$  For both the study sites
3: Outputs: Trained model with R, RMSE and bias value  $\triangleright$  For both the study sites
4: procedure GENERATEFISUSINGSHS(data)
5:   ClusNum  $\leftarrow$  5 and MaxIt  $\leftarrow$  1000  $\triangleright$  Number of clusters in FCM and maximum iterations
6:   fis  $\leftarrow$  GenerateFuzzy(data, ClusNum)  $\triangleright$  Generate Fuzzy Model
7:   SHSFis  $\leftarrow$  shs(fis, data)  $\triangleright$  Training SHS Algorithm
8:   Return SHSFis
9: end procedure
10: procedure GENERATEFUZZY(data, ClusNum)  $\triangleright$  Generate Fuzzy Model
11:   if  $\neg \text{exist}(\{ClusNum', \}var')$  then
12:     ClusNum  $\leftarrow$  auto'
13:   end if
14:   x  $\leftarrow$  data.TrainInputs; t  $\leftarrow$  data.TrainTargets
15:   Calculate membership values using Gaussian membership function (gaussmf)  $\triangleright$  Initialise cluster centers randomly
16:   fcm_U  $\leftarrow$  2; fcm_MaxIter  $\leftarrow$  100; fcm_MinImp  $\leftarrow$   $1e-5$   $\triangleright$  FCM parameters
17:   fcm_options  $\leftarrow$  [fcm_U fcm_MaxIter fcm_MinImp fcm_Display]
18:   Linguistic_variables = DefineLinguisticVariables()  $\triangleright$  Define linguistic variables
19:   Fuzzy_rules = GenerateFuzzyRules(Linguistic_variables)
20:   Return fis
21: end procedure
22: procedure SHS(fis, data)
23:   p0  $\leftarrow$  GettingFuzzyParameters(fis)
24:   CostFunction  $\leftarrow$   $\lambda x$  : FuzzyCost(x, fis, data)
25:   nVar  $\leftarrow$  numel(p0)
26:    $\alpha \leftarrow 1$ ; VarMin  $\leftarrow$   $-(10^\alpha)$ ; VarMax  $\leftarrow 10^\alpha$   $\triangleright$  Decision variables lower and upper bound
27:   nPop  $\leftarrow$  10  $\triangleright$  Population size
28:    $\omega \leftarrow \text{randomInteger}(1, 3)$  and  $\psi \leftarrow \text{randomInteger}(1, 3)$   $\triangleright$  Claw power ( $\omega$ ) and Sting power ( $\psi$ ) coefficient
29:    $\mu \leftarrow 0.2$  and  $\mu_{damp} \leftarrow 0.98$   $\triangleright$  Mutation rate ( $\mu$ ) and Mutation coefficient damping ratio ( $\mu_{damp}$ )
30:    $\delta \leftarrow 0.05 \times (\text{VarMax} - \text{VarMin})$   $\triangleright$  Uniform mutation range
31:   for it = 1 to MaxIt do  $\triangleright$  Initialise population and best solution
32:     Update population using SHS algorithm
33:     Damp mutation coefficient
34:   end for
35:   Return results
36: end procedure
37: procedure MAIN
38:   for each data in data.Test do
39:     model_SM = SHSFis(data.Test)
40:   end for
41:   R, RMSE, bias = computeMetrics(in-situ_SM, model_SM)
42: end procedure
43: Main()

```

91 ground-based observations are collected during field campaigns on January 17 and 29, 2019, and October 16 and 28, 2022. For the Kosi Fan region, 281 measurements are obtained across two field seasons: December 11, 15, 17, 18, and 20, 2019, and March 1, 3, and 6, 2022. A linear regression analysis is performed between the downscaled estimates and corresponding in-situ values to assess the model's predictive skill.

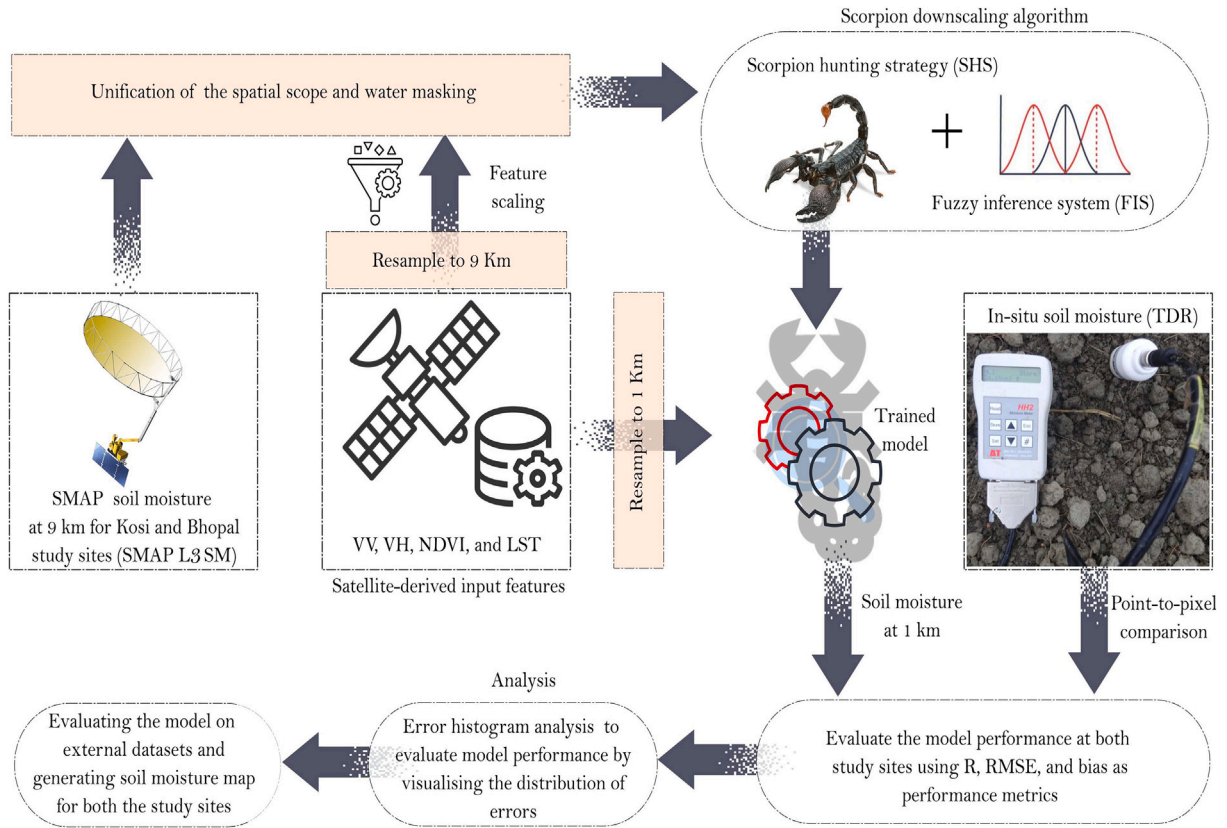


Fig. 5. Outline of the downscaling process adopted in this study.

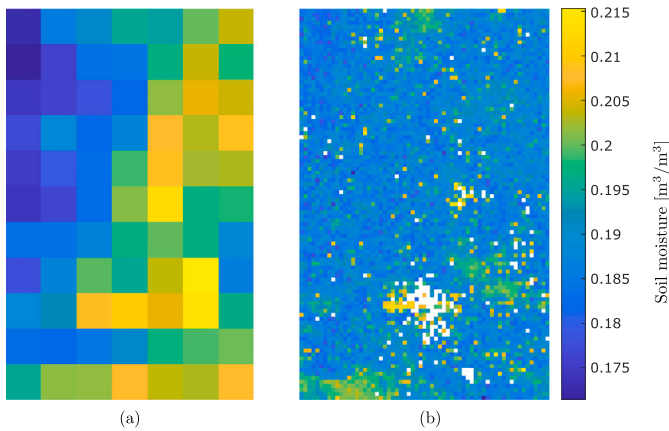


Fig. 6. (a) Soil moisture on January 17, 2019, derived from SMAP data at 9 km resolution. (b) Downscaled soil moisture map for January 17, 2019, derived from the scorpion downscaling algorithm at a spatial resolution of 1 km.

At the Bhopal site, the downscaled soil moisture exhibited strong agreement with field observations, with a Pearson correlation coefficient (R) of 0.79, root mean square error (RMSE) of $0.053 \text{ m}^3/\text{m}^3$, and a slight negative bias of -0.019 (Fig. 8(a)). An error histogram analysis (Fig. 8(b)), fitted with a Gaussian curve across seven bins, shows a distribution centered near the zero-error line. This indicates that most prediction errors are minor and symmetrically distributed. This confirms the model's strong fit and minimal systematic error in a semi-arid setting.

Similarly, on the Kosi Fan (Fig. 9), the downscaled estimates showed a correlation of $R = 0.76$, RMSE of $0.043 \text{ m}^3/\text{m}^3$, and a slight positive bias of 0.037. The Gaussian fit of the error histogram also peaked

near zero, further validating the model's capacity to accurately estimate soil moisture in a humid sub-tropical environment. The lower RMSE in this region highlights the algorithm's ability to generalise well even in landscapes with dense vegetation and persistent surface moisture.

The similarity in performance across two ecologically diverse regions confirms the downscaling method's flexibility and generalisation. The algorithm successfully mimicked sharp spatial gradients of soil moisture in an environment of low moisture, despite inherent challenges such as high evaporation and fragmented land cover in Bhopal. The slightly negative bias proves to be a conservative estimate and is still within operationally acceptable ranges.

In contrast, the Kosi Fan region, characterised by higher baseline soil moisture and more homogeneous land surface conditions, showed a small overestimation bias. This is likely attributable to persistent saturated soils influencing the input features. Nevertheless, the overall low error metrics confirm the model's suitability for use in hydrologically active zones.

In conclusion, the downscaling framework proved effective in producing high-resolution soil moisture estimates that align closely with ground-truth observations, while also enhancing spatial detail well beyond the native SMAP resolution. This capability is critical for a range of applications, such as precision agriculture, drought monitoring, and hydrological modeling, where timely and spatially explicit soil moisture information is essential for informed decision-making.

4.2. Comparison with previous studies

The proposed downscaling algorithm demonstrates competitive performance compared to the recent studies (Table 2). Our approach, based on the Scorpion Downscaling Algorithm (SDA), achieves correlation coefficients R ranging from 0.76 to 0.79, surpassing several existing methodologies, including those using Thermal Inertia Theory (TIT) and

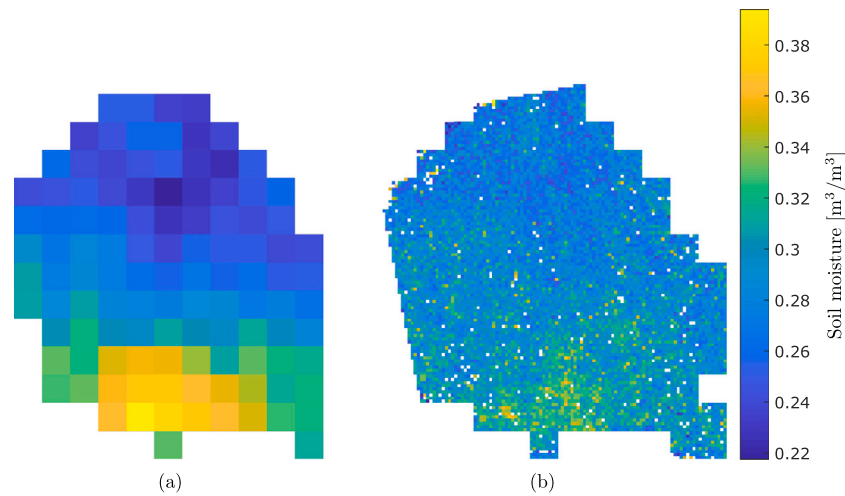


Fig. 7. (a) Soil moisture on December 15, 2019, derived from SMAP data at 9 km resolution. (b) Downscaled soil moisture map for December 15, 2019, derived from the scorpion downscaling algorithm at a spatial resolution of 1 km.

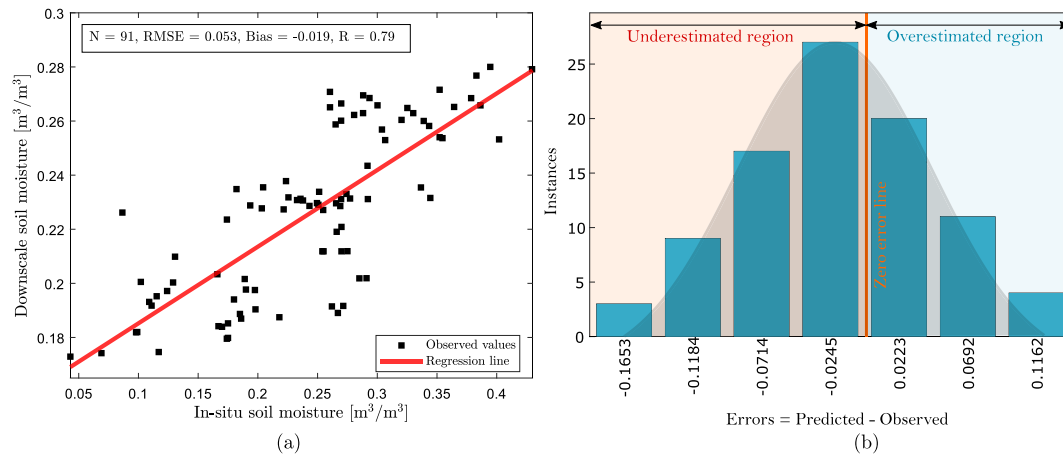


Fig. 8. (a) Linear fit between the downscaled and in-situ soil moisture for Bhopal. (b) Error analysis using histogram with seven bins depicting the distribution of errors in the scorpion downscaling compared to in-situ observations. The red line represents the zero error line.

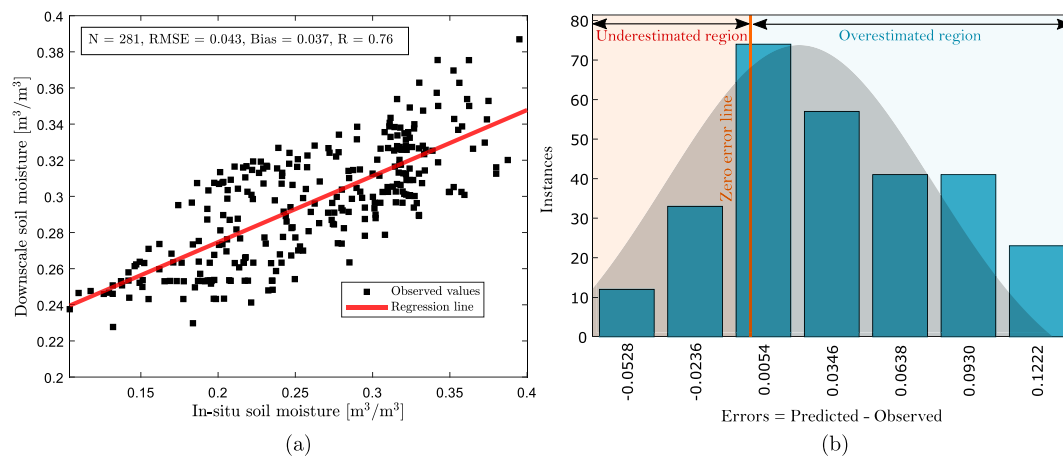


Fig. 9. (a) Linear fit between the downscaled and in-situ soil moisture for the Kosi Fan. (b) Error analysis using histogram with seven bins depicting the distribution of errors in the scorpion downscaling compared to in-situ observations. The red line represents the zero error line.

Table 2
Comparison with recent advanced studies in downscaling SMAP soil moisture products.

Publications	[22]	[7]	[23]	[24]	[25]	[26]	[27]	[28]	This study
Best algorithm	TTT	CNN	TTT	TTT	NN	DBN	RF	DISPATCH	SDA
Study site	Walnut Gulch	Oklahoma	Mekong	Oklahoma	US (CONUS)	SRB	LP	SF	KFB
Performance metrics									
R	0.30 to 0.86	0.659	0.435 to 0.836	0.239 to 0.754	0.715	0.725	0.654	0.36 to 0.79	0.76 to 0.79
RMSE [m^3/m^3]	–	8.40%	–	–	–	0.0303	–	–	0.043 to 0.053
ubRMSE [m^3/m^3]	0.006 to 0.088	5.18%	–	0.036 to 0.079	0.041	–	0.044	–	–
Bias [m^3/m^3]	–0.079 to –0.002	–	–	0.033 to 0.056	–	–	–	–	–0.019 to 0.037

Note: TTT stands for Thermal inertia theory, NN stands for Neural networks, SDA stands for Scorpion downscaling algorithm, SF stands for Spain and France, DBN stands for Deep Belief Network, SRB stands for Shandian River Basin, KFB stands for Kosi Fan and Bhopal, LP stands for Loess Plateau.

Table 3

Performance comparison of the proposed algorithm across four publicly available datasets using different model configurations. Metrics are reported separately for SHS and Primary subsets.

Performance metrics	Model variants and datasets									
	AutoML-ID [29]		LT-FS-ID [30]		FF-ANN-ID (Gaussian Init.) [31]		FF-ANN-ID (Uniform Init.) [31]		ALE [32]	
	SHS	Primary	SHS	Primary	SHS	Primary	SHS	Primary	SHS	Primary
R	0.99	1.00	0.99	0.98	0.99	0.78	0.82	0.79	0.71	0.82
RMSE	3.36	0.01	0.01	6.47	5.97	41.45	27.40	48.36	0.14	0.15
Bias	0.99	–0.01	0.00	12.35	1.07	3.02	–5.92	9.55	0.06	0.05

Neural Networks (NN), which reported values as low as 0.239. In terms of RMSE, our results (0.043 to $0.053 \text{ m}^3/\text{m}^3$) are comparable to the Deep Belief Network (DBN)-based method ($0.0303 \text{ m}^3/\text{m}^3$), while outperforming studies where RMSE exceeded 8.40%. Additionally, the RMSE values in our study align closely with those reported in previous works, such as Fang et al. (2018) and Wang et al. (2024). The bias, ranging from -0.019 to $0.037 \text{ m}^3/\text{m}^3$, indicates minimal systematic errors compared to earlier downscaling methods. These results highlight the effectiveness of SDA in improving soil moisture downscaling, particularly for regions with contrasting hydrological characteristics, making it a promising nature-inspired approach for future applications.

4.3. Performance over publicly available datasets

We evaluated the performance of the scorpion downscaling algorithm as a generalised regression tool. We employed publicly available datasets from databases such as UCI and Kaggle, i.e., AutoML-ID [29], LT-FS-ID [30] (<https://archive.ics.uci.edu/dataset/715/lt+fs+id+intrusion+detection+in+wsns>), FF-ANN-ID [31] (<https://www.kaggle.com/datasets/abhilashdata/ffannid-intrusion-detection-in-wsns>), and ALE [32] ([https://archive.ics.uci.edu/dataset/844/average+localization+error+\(ale\)+in+sensor%20+node+localization+process+in+wsns](https://archive.ics.uci.edu/dataset/844/average+localization+error+(ale)+in+sensor%20+node+localization+process+in+wsns)). Each dataset was divided randomly into training and test subsets using the Mersenne Twister random generator, with a ratio of 70:30. The algorithm was trained on 70% of the data and its performance was tested on the remaining 30% using R, RMSE, and bias as performance measures (Table 3).

We found that the scorpion downscaling algorithm is good as a generalised regression method, providing excellent performance in all four datasets. We also compared its performance with the main algorithms on the same datasets. Interestingly, in the case of the LT-FS-ID and FF-ANN-ID datasets, the proposed algorithm outperformed the main algorithms, establishing its greater efficacy.

The extensive applicability of the scorpion algorithm for downscaling extends beyond its first application in downscaling SMAP soil moisture from 9 km to 1 km. Its consistency across diverse datasets speaks volumes about its potential for use in an extensive variety of fields such as environmental modeling, intrusion detection analysis, financial

forecasting, and many more. That the algorithm generalises so effectively to different types of data is a testament to its versatility and applicability to researchers and practitioners in a broad array of disciplines. By resolving challenging regression issues, the scorpion downscaling algorithm unlocks the potential for advancement in data-driven decision making, along with predictive modeling in a broad array of applications.

4.4. Limitations and future work

Despite the encouraging results, the scorpion downscaling algorithm does have some disadvantages. Firstly, the performance of the algorithm will be affected by the quality and resolution of input data. Errors and inconsistencies within the input features can potentially deteriorate the downscaled output. In addition, even though the algorithm has done reasonably well with sub-tropical dry and humid sub-tropical areas, how efficiently it can cope with other climatic zones such as arid deserts or tropical rainforests remains to be thoroughly tested. Another weakness lies in the computational complexity of nature-inspired optimisation and fuzzy inference systems, which would lead to long processing times with large amounts of data. Further, the algorithm might require parameter tuning for different sets of data in order to get the optimum performance, and that can be cumbersome and domain-skill intensive. Finally, the risk of overfitting in regions of sparse ground-truth measurements justifies rigorous validation and cross-validation techniques in order to be certain about the model's generalisability.

Given the strong cross-domain generalisation observed across multiple public datasets, a promising future direction is to adapt the SHS-FIS framework for precipitation downscaling. In particular, extending the method to map coarse-resolution reanalysis products such as ERA5 to IMERG-like fine spatial resolutions remains an open and impactful challenge. The interpretable fuzzy rule structure paired with SHS optimisation makes the approach well-suited for capturing multiscale precipitation patterns, and this will be explored in future work.

Another important future avenue is to evaluate the transferability of the proposed algorithm across different geographic regions using transfer learning. For example, an SHS-FIS model pre-trained on a data-rich location such as the Kosi Fan could be fine-tuned on a data-scarce

target region like the Bhopal using limited in-situ points. Strategies such as (i) freezing the fuzzy rule base and adapting only the consequent parameters, (ii) warm-starting the SHS optimiser, or (iii) few-shot fine-tuning can be tested. Such experiments would quantify cross-site generalisation and support practical deployment in new locations with minimal ground data.

5. Conclusion

This article proposes a novel scorpion downscaling algorithm by combining the scorpion hunting strategy algorithm (a bio-inspired algorithm) with a fuzzy inference system to downscale SMAP soil moisture data from 9 km to 1 km. We rigorously compared the performance using in-situ measurements of two diverse climatic zones: Bhopal (semi-arid) and Kosi Fan (humid sub-tropical) which demonstrate the strong performance and flexibility of the algorithm. Compared with widely used meta-heuristic optimisers in remote sensing, such as PSO, DE, and GA, the Scorpion Hunting Strategy (SHS) provides a more adaptive exploration-exploitation balance through its biologically inspired α , β , and α - β mutation operators. This dynamic behaviour enables SHS to navigate nonlinear and heterogeneous downscaling landscapes more effectively, allowing it to capture fine-scale soil moisture patterns with greater robustness.

Downscaled soil moisture maps confirm outstanding agreement with ground truth data, indicating high precision in capturing intricate spatial variations. The low error statistics and small biases also validate the reliability of the algorithm.

These results validate the capability of the scorpion downscaling algorithm to serve as a universal regression instrument in providing precise soil moisture data under various physical and environmental conditions. By integrating an optimisation strategy inspired by nature and fuzzy logic, our approach represents a major milestone in acquiring nature-inspired solutions to environmental modelling. The applicability of this algorithm is wide-ranging beyond the downscaling of soil moisture and can be a valuable source in obtaining nature-based solutions for diverse applications in hydrology, agriculture, and climate studies.

CRedit authorship contribution statement

Abhilash Singh: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **M. Niranjannaik:** Data curation, Validation, Writing – original draft preparation, Visualisation, Investigation. **Kumar Gaurav:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Data curation.

Software availability

- Software name: NiSSMDA (Nature-inspired Spatial Soil Moisture Downscaling Algorithm)
- Developer: Singh et al.
- Contact information: a.singh4@leeds.ac.uk
- First year available: 2026
- Programming language: MATLAB
- Cost: Free
- Software availability: <https://abhilashsingh.net/codes.html>

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This research was supported by the Space Applications Centre, Indian Space Research Organisation (SAC-ISRO), as part of the NASA-ISRO Synthetic Aperture Radar (NISAR) mission, under Grant HYD-01.

Data availability

- SMAP data: <https://nsidc.org/data/smap/data>
- Input features:
Sentinel-1/2 <https://scihub.copernicus.eu/> or <https://browser.dataspace.copernicus.eu/>
MODIS LST: <https://modis.gsfc.nasa.gov/>
- Response variable: The in-situ data used in this study are available in our previous publications [13].
- Infrastructure used: NVIDIA A40 GPU with 46GB VRAM and an Intel Xeon Gold 6338 CPU (64 cores, 128 threads)

References

- [1] Q. Wang, R. van der Velde, P. Ferrazzoli, X. Chen, X. Bai, Z. Su, Mapping soil moisture across the Tibetan plateau plains using aquarius active and passive l-band microwave observations, *Int. J. Appl. Earth Obs. Geoinf.* 77 (2019) 108–118.
- [2] M. Schönauer, R. Prinz, K. Väättäinen, R. Astrup, D. Pszenny, H. Lindeman, D. Jaeger, Spatio-temporal prediction of soil moisture using soil maps, topographic indices and smap retrievals, *Int. J. Appl. Earth Obs. Geoinf.* 108 (2022) 102730.
- [3] J. Bai, Q. Cui, W. Zhang, L. Meng, An approach for downscaling smap soil moisture by combining sentinel-1 SAR and modis data, *Remote Sens.* 11 (23) (2019) 2736.
- [4] P.M. Atkinson, Downscaling in remote sensing, *Int. J. Appl. Earth Obs. Geoinf.* 22 (2013) 106–114.
- [5] W. Zhao, N. Sánchez, H. Lu, A. Li, A spatial downscaling approach for the smap passive surface soil moisture product using random forest regression, *J. Hydrol.* 563 (2018) 1009–1024.
- [6] H. Zhang, B. Huang, Support vector regression-based downscaling for intercalibration of multiresolution satellite images, *IEEE Trans. Geosci. Remote Sens.* 51 (3) (2013) 1114–1123.
- [7] W. Xu, Z. Zhang, Z. Long, Q. Qin, Downscaling smap soil moisture products with convolutional neural network, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 14 (2021) 4051–4062.
- [8] A. Kumar, K. Zhang, Improving total phosphorus prediction and monitoring in data scarce streams using a hybrid machine learning framework, *J. Hydrol.* (2025) 134636.
- [9] A. Singh, S.M.H. Mousavi, K. Gaurav, Shs: Scorpion hunting strategy swarm algorithm, *arXiv preprint arXiv:2407.14202*, 2024.
- [10] NRSC, District and category wise distribution of land use and land cover in Madhya Pradesh (2015–16), <https://bhuvan-app1.nrsc.gov.in/2dresources/thematic/LULC503/MAP/MP.pdf> (accessed 25 May 2024).
- [11] A. Singh, M.N. Naik, K. Gaurav, Drainage congestion due to road network on the kosi alluvial fan, himalayan foreland, *Int. J. Appl. Earth Obs. Geoinf.* 112 (2022) 102892.
- [12] NRSC Bhuvan, District and category wise distribution of land use and land cover in Bihar (2015–16), <https://bhuvan-app1.nrsc.gov.in/2dresources/thematic/LULC503/MAP/BR.pdf> (accessed 25 March 2024).
- [13] A. Singh, K. Gaurav, Piml-sm: physics-informed machine learning to estimate surface soil moisture from multi-sensor satellite images by leveraging swarm intelligence, *IEEE Trans. Geosci. Remote Sens.* 62 (2024) 4416913, <https://doi.org/10.1109/TGRS.2024.3502618>
- [14] Z. Xing, X. Li, L. Fan, A. Colliander, F. Frappart, P. de Rosnay, R. Fernandez-Moran, X. Liu, H. Wang, L. Zhao, et al., Assessment of 9 km smap soil moisture: evidence of narrowing the gap between satellite retrievals and model-based reanalysis, *Remote Sens. Environ.* 296 (2023) 113721.
- [15] D. Entekhabi, S. Yueh, P.E. O'Neill, K.H. Kellogg, A. Allen, R. Bindlish, M. Brown, S. Chan, A. Colliander, W.T. Crow, et al., Smap handbook–soil moisture active passive: mapping soil moisture and freeze/thaw from space, https://smap.jpl.nasa.gov/files/smap2/SMAP_handbook_web.pdf, 2014.
- [16] P. O'Neill, S. Chan, E. Njoku, T. Jackson, R. Bindlish, J. Chubb, A. Colliander, Smap enhanced l3 radiometer global and polar grid daily 9 km ease-grid soil moisture, version 5, https://nsidc.org/data/spl3smp_e/versions/5/, 2021.
- [17] F. Wen, W. Zhao, Q. Wang, N. Sánchez, A value-consistent method for downscaling smap passive soil moisture with modis products using self-adaptive window, *IEEE Trans. Geosci. Remote Sens.* 58 (2) (2019) 913–924.
- [18] M. Sugeno, *Industrial Applications of Fuzzy Control*, Elsevier Science Inc., 1985.
- [19] R. Meyer, W. Zhang, S.J. Kragh, M. Andreasen, K.H. Jensen, R. Fensholt, S. Stisen, M.C. Looms, Exploring the combined use of smap and sentinel-1 data for downscaling soil moisture beyond the 1 km scale, *Hydrol. Earth Syst. Sci.* 26 (13) (2022) 3337–3357.
- [20] E.P. Maurer, H.G. Hidalgo, Utility of daily VS. Monthly large-scale climate data: an intercomparison of two statistical downscaling methods, *Hydrol. Earth Syst. Sci.* 12 (2) (2008) 551–563.
- [21] T. Vandal, E. Kodra, A.R. Ganguly, Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation, *Theor. Appl. Climatol.* 137 (2019) 557–570.
- [22] B. Fang, V. Lakshmi, R. Bindlish, T.J. Jackson, Downscaling of smap soil moisture using land surface temperature and vegetation data, *Vadose Zone J.* 17 (1) (2018) 1–15.
- [23] C. Dandridge, B. Fang, V. Lakshmi, Downscaling of smap soil moisture in the lower Mekong river basin, *Water* 12 (1) (2019) 56.

- [24] B. Fang, V. Lakshmi, R. Bindlish, T.J. Jackson, Amsr2 soil moisture downscaling using temperature and vegetation data, *Remote Sens.* 10 (10) (2018) 1575.
- [25] M. Xu, N. Yao, H. Yang, J. Xu, A. Hu, L.G.G. de Goncalves, G. Liu, Downscaling smap soil moisture using a wide & deep learning method over the continental united states, *J. Hydrol.* 609 (2022) 127784.
- [26] Y. Cai, P. Fan, S. Lang, M. Li, Y. Muhammad, A. Liu, Downscaling of smap soil moisture data by using a deep belief network, *Remote Sens.* 14 (22) (2022) 5681.
- [27] Y. Wang, H. Shi, X. Yang, Y. Jiang, Y. Wu, J. Shui, Y. Liu, M. Guo, L. Li, Spatial downscaling of smap soil moisture to high resolution using machine learning over china's loess plateau, *Catena* 247 (2024) 108492.
- [28] N. Ojha, O. Merlin, C. Suere, M.J. Escorihuela, Extending the spatio-temporal applicability of dispatch soil moisture downscaling algorithm: a study case using smap, modis and sentinel-3 data, *Front. Environ. Sci.* 9 (2021) 555216.
- [29] A. Singh, J. Amutha, J. Nagar, S. Sharma, C.C. Lee, Automl-id: automated machine learning model for intrusion detection using wireless sensor network, *Sci. Rep.* 12 (1) (2022) 1–14.
- [30] A. Singh, J. Amutha, J. Nagar, S. Sharma, C.-C. Lee, Lt-fs-id: log-transformed feature learning and feature-scaling-based machine learning algorithms to predict the k-barriers for intrusion detection using wireless sensor network, *Sensors* 22 (3) (2022) 1070.
- [31] A. Singh, J. Amutha, J. Nagar, S. Sharma, A deep learning approach to predict the number of k-barriers for intrusion detection over a circular region using wireless sensor networks, *Expert Syst. Appl.* (2022) 118588.
- [32] A. Singh, V. Kotiyal, S. Sharma, J. Nagar, C.-C. Lee, A machine learning approach to predict the average localization error with applications to wireless sensor networks, *IEEE Access* 8 (2020) 208253–208263.